DOCUMENT RESUME

ED 432 263 IR 019 633

AUTHOR Lansford, Carl E.
TITLE Using Pre-test/Post-test Data To Evaluate the Effectiveness of Computer Aided Instruction (A Study of CAI and Its Use with Developmental Reading Students).
PUB DATE 1999-03-00
NOTE 7p.; In: SITE 99: Society for Information Technology & Teacher Education International Conference (10th, San Antonio, TX, February 28-March 4, 1999); see IR 019 584.
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Computer Assisted Instruction; Distance Education; Evaluation Methods; Higher Education; *Instructional Effectiveness; Instructional Material Evaluation; Pretests Posttests; Program Effectiveness; *Research Design
IDENTIFIERS Texas Southmost College; University of Texas Brownsville

ABSTRACT

As computer aided instruction (CAI) and distance learning become more popular, a model for easily evaluating these teaching methods must be developed, one which will enable replication of the study each year. This paper discusses the results of a study using existing dependent and independent variables to evaluate CAI for developmental reading students at Texas Southmost College and the University of Texas at Brownsville (UTB/TSC). Pre-test and post-test data, time-on-task in CAI, and multiple regression analysis were used. The following five threats to the internal validity of the one-group, pre-test-posttest design are explained: history, maturation, testing, instrumentation, and interaction of selection and any one of the other threats to internal validity. (AEF)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 432 263

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Using Pre-test/Post-test Data to Evaluate the Effectiveness of Computer Aided Instruction

(A study of CAI and its use with Developmental Reading students)

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

G.H. Marks

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Carl E. Lansford
Academic Computing Department
The University of Texas at Brownsville and Texas Southmost College, USA
carl@utb.edu

Abstract: As Computer Aided Instruction and Distance Learning become more popular, a model for easily evaluating these teaching methods must be developed, one which will enable replication of the study each year. This paper discusses the results of a study using existing pretest and posttest data to evaluate CAI for developmental reading students.

Introduction

Texas Southmost College in partnership with The University of Texas at Brownsville (UTB/TSC) serves approximately 10,000 students in a large portion of the Lower Rio Grande Valley in the southern tip of Texas. Because of the open enrollment policy at TSC, many UTB/TSC students enter college with deficiencies in reading ability, in mathematics competency, or in writing skills.

Beginning in 1988, the state of Texas mandated the Texas Academic Skills Program (TASP) for all new college students. If a student does not pass the reading, mathematics and writing portions of the TASP he/she must take developmental courses until each portion of the TASP is passed. Approximately twenty-percent of the 62,731 semester hours taken by TSC students each semester are in developmental courses (20.1% in Spring, 1997).

The cost of developmental reading courses is high. Classes must be small so the teacher will be more available to the students. Teachers must be trained in developmental reading. Every faculty member was teaching several developmental reading classes, and many adjunct-faculty also were hired. Even with small classes, the success rate of developmental reading students was low. UTB/TSC has been seeking ways to improve the success of students in developmental courses, but also to lower the costs of developmental training.

Computer Aided Instruction can be justified on a financial basis in today's world. After an initial investment, a large number of students can be served on a continuing basis with CAI. Some studies have shown that CAI is successful, but other studies have shown that CAI is not always successful. There are reasons to believe CAI will be a satisfactory substitute for traditional developmental courses at UTB/TSC:

- Students working in CAI will be able to work at convenient times.
- Students taking developmental reading by Distance Education will be able to work at convenient times without the necessity of commuting.
- In CAI students are pre-tested and a personalized course of study is prescripted for each student.
- In CAI a student can work at his own pace.
- In CAI students will spend their time on skills they need to learn.
- In CAI students will have tutorials, practice exercises, and tests over each skill.

TASP exam results for CAI students have shown that CAI is approximately as successful as the previous results from instructor-led courses, but comparisons between the two methods have been disappointing. CAI has been allowed to continue, mostly because some of the problems with the computer lab are solved each semester.

CAI has proven to be a cost-effective way of providing developmental reading. Students enter developmental reading at many different levels, so traditional teaching of developmental courses was a

IR019633

BEST COPY AVAILABLE

complex and cumbersome task. Each student in CAI is scheduled for three hours weekly in a large computer lab. Peer tutors and faculty are available if students need help. More than 100 computers are busy most of a 14-hour day. Since the computer administers placement tests, and prescribes which modules a student will need to complete, the CAI Lab is called Computer Directed Instruction (CDI).

Need for the study

A study by the National Center for Developmental Education (NCDE, 1996) found that the quality of remedial courses provided in support of the TASP test varies widely from college to college. The study stated:

> Ongoing and systematic evaluation of the outcomes of remedial courses and programs is rare. Although there is a strong emphasis on compliance with TASP regulations among Texas colleges and universities there is little emphasis on accountability for the outcomes of TASP remediation" (Boylan, 1998).

As CAI becomes more popular, a method for studying the effectiveness of CAI must be developed, one which will allow replication of the study each year.

Problems with using an experimental model for evaluation

A randomized experimental model with two groups, if it were possible, would not provide causal inferences that meet the needs of the policy planners. Next semester's students are not part of the population studied, and there is no scientific justification for applying causal inferences to them. Only after several semesters of replication, could common inferences be applied to similar populations (as long as the labs are similar and the students are using the same CDI software).

True experimental studies in an educational setting are seldom acceptable. The experimental process would require the researcher to schedule random samples from the entire population of developmental reading students and to place them into control and experimental groups. The researcher would have to control all factors affecting the student's schedule (work, other classes, family responsibilities, commuting, etc.).

There are multiple variables affecting these developmental students in their first semesters at college. All of these variables affect their progress in developmental reading. Achen describes the idiosyncratic behavior as a function of the history and personality of the humans studied:

> Any realistic data set involves a hopeless jumble of human actors, all engaging in idiosyncratic behavior as a function of numberless distinctive features of their histories and personalities. Many thousands of details of their individual histories contribute to their behavior (Achen, 1992, p. 25).

Achen continues, "Functionally correct causal specification in social science is neither possible nor desirable" (Achen, 1992, p. 25).

Problems with using a quasi-experimental model for evaluation

Quasi-experimental studies are commonly used in the educational setting. In a quasi-experimental model, students are allowed to choose their classes, and then some classes are assigned to the control group and others to the experimental group. Because no random assignments are made in quasi-experimental research, the samples are not representative of any population. After a quasi-experimental study, some combination of other factors, such as pupil or teacher satisfaction (or cost), are usually relied on to determine whether or not to continue using CAI. Indeed, a quasi-experimental study cannot result in a determination that CAI provides the best teaching method for teaching future students, or even for teaching the current population. The benefits of the two-group method used in the quasi-experimental mode are overshadowed by the dynamic nature of computer software and CAI instruction.

The developmental reading computer lab at UTB/TSC has gone through many changes and has been expanded. The DOS-based PLATO software, developed in the 1960's by Control Data Corporation and the University of Illinois, was expanded by upgrades, and improvements were added several times during the two years of use. A new Windows-based software package, "Destinations" was selected in 1996. A brief summary of events during this study shows the fallacy of applying causal inferences to future CAI labs:

- First semester--too few computers, long lines of students unable to get computer time.
- Second semester--better scheduling of computer time but no way to enforce time-on-task.
- Third semester—better scheduling of computer use, but the lab suffered for a month because a hacker kept erasing student data files.
- Fourth semester--few problems this semester, but no faculty involvement and no peer tutors.
- Fifth semester--entirely new software package, new lab, better scheduling.
- Sixth semester--more peer tutors and teachers hired.
- Seventh semester--major upgrade in the CAI software, network problems resulted in rewiring the lab network.
- Eighth semester--less problems but more software upgrades.
- In 1999, upgrade to a new Internet-based software delivery.

Software upgrades, the addition of peer tutors, changes in scheduling methods, and many other factors kept changing each semester. Which of the changes were effective and which were not? UTB/TSC needs an evaluation model that can be used to *continually* evaluate the success of CDI. The count of the number of students completing the reading portion of the TASP exam does not provide enough information to properly evaluate the effectiveness of a required program which accounts for approximately twenty-percent of UTB/TSC student's total semester hours.

The costly control-group/experimental-group would have to be repeated each semester to provide needed information for administrators. Results from such research could not justifiably be applied to make inferences about the next semester.

Determining a research model is the first step in a deeper study of the effectiveness of CAI/CDI as a teaching tool for developmental reading. According to Fox (1997), "Statistical models are capable of capturing and describing that structure, or at least significant aspects of it." (Fox, 1997, p. 5)

Using a statistical control model for evaluation

For the purpose of this study correlation does not provide enough information about the relationship between variables. It is not enough to say that a correlation exists between the variables being studied. As Elizabeth Cohen describes the problem: "We need to construct a more powerful statement for an ideal policy proposition, one using something similar to the following form: if we vary x, then certain predictable changes will occur in y." She continues, "This proposition requires no less than an effective understanding of how the two variables come to be associated" (E. Cohen, 1970, 28-29).

Further, changes (hopefully, "improvements"), were made each semester in the way CAI and then CDI were applied: more computers, more tutors, better software, and many other "improvements" were made each semester.

Statistical control can provide a means for studying the success of CAI and distance learning without creating an artificial environment for developmental students. Statistical control and its value is described by Pedhazur in the following paragraph:

> Statistical control means that one uses statistical methods to identify, isolate, or nullify variance in a dependent variable that is presumably "caused by one or more independent variables that are extraneous to the particular relation or relations under study. Statistical control is particularly important when one is interested in the joint or mutual effects of more than one independent variable on a dependent variable because one has to be able to sort out and control the effects of some variables while studying the effects of other variables. Multiple regression and related forms of analysis provide ways to achieve such control (Pedhazur, 1982, p. 98).

Since all students take a placement exam before they enroll in developmental reading and the TASP exam after they have taken developmental reading, these measures can be statistically adjusted to

provide pretest and posttest observations. The CAI software keeps track of the time that students are working exercises in CAI. These variables can be used to answer the question, "Does a reasonable amount of time-on-task in CAI prepare developmental reading students for successful completion of the TASP exam?" The following variables can be used to develop a statistical model for evaluation of CAI.

Dependent variables--*pretest* and *posttest*: In this study, reading score on the TASP exam is the posttest measure. Pretest measures were either the Nelson-Denny Reading Test Form E or the ACT-ASSET test, which replaced the Nelson-Denny in the Fall of 1995.

Independent variable--*time-on task in CAI*: Time-on-task in the CAI experience is used as the experimental variable. The variable used in this study will be the measure of time the student spends working in the CAI reading program. Although many other school experiences help the student learn to read, students spending more time using the CAI program will also receive more benefit from other school experiences. Time-on-task in CAI will be used as a measure of the "CAI experience".

The research model for this study will be a single-group, pretest/posttest model using multiple regression to assess the relationship between the time-on-task in the CAI experience and progress in reading as shown on the TASP exam. The results are determined by comparing the pretest score to the posttest score. This study will be replicated for a period of eight semesters using eight populations and two different CAI programs to determine if results are consistent. Results from this study will be described in this presentation. The research hypothesis is *"Students who spend more time-on-task in CAI will make better score gains when they take the TASP exam than students who spend less time-on-task in CAI."*

Why use multiple regression?

Multiple regression provides a method for statistically controlling for multiple factors affecting reading progress, and for developing predictive equations useful in evaluation of CAI. In this study, multiple regression analysis is used to provide a scientific explanation of something that has happened, rather than to forecast the future. Although causal inferences cannot necessarily be developed from a regression analysis, such an analysis does make possible a meaningful statistical interpretation of the relationship between variables

Kleinbaum and Kupper state that MRC (Multiple Regression/Correlation) can be used, "To describe the extent, direction, and strength of the relationship between several independent variables and a continuous dependent variable" (Kleinbaum & Kupper, 1978, p. 11). They continue, stating that multiple regression analysis is a general technique, which can be used with all kinds of variables (Kleinbaum & Kupper, 1978, p. 14). Cowen and Cowen describe the process,

> The basic strategy of the analysis of causal models is first to state a theory in terms of the variables that are involved and, quite explicitly, of what causes what and what does not, usually aided by causal diagrams. The observational data are then employed to determine whether the causal model is consistent with them, and estimate the strength of the causal parameters (Cohen & Cohen, 1983, p. 14).

Limitations of the study

The one-group, pretest-posttest design chosen for this study has several limitations that must be considered. The pretest-posttest is used to rule out selection as a rival explanation, but Kidder lists the five other threats to the internal validity of the one-group, pretest-posttest design. The threats are (1) history, (2) maturation, (3) testing, (4) instrumentation, and (5) interaction of selection and maturation (Kidder, 1970, p. 45). These five rival explanations are potential threats for any study using the one-group, pretest-posttest design and they will be ruled out, in part, by limiting the conclusions from this study. Each of the rival explanations is described below.

History. The posttest is not taken until several months after the pretest (perhaps even a year or two later) so part of the difference between the posttest and the pretest may result from the different social climate. The student is now a college student, pressured to study (and to read effectively) by faculty in other

courses. No meaningful study of college reading effectiveness would want to isolate the student from the rival explanation of history. Such a study would require an artificial environment for the student, and could test only some small part of CAI instruction.

This study seeks to include the effects of history as a part of the student's success in developing reading ability (as judged by the TASP exams). The study does not seek to isolate CAI as a separate variable, but will use as a variable "the CAI experience", which includes CAI and college experiences (history) during the time the student takes developmental reading. History is not a rival explanation, but is a part of the variable being studied.

Maturation. Certainly new college students are maturing during the first year or two of college. They have become independent in many ways after leaving high school--some are working, some are married, and most have less family control over them. They have become responsible for their own learning like never before. Maturation is an important factor in college success. Like history, maturation becomes a part of the variable being studied in "the CAI experience" and the researcher does not seek to isolate it for the purposes of this study. Maturation is not a rival explanation, but is a part of the variable being studied.

Testing. The tests taken in this study are required for every student entering college. Certainly the pretest will affect the developmental student because he failed to make an acceptable score on the test. The pretest shows the student that he needs to acquire additional reading skills. Tests are used in a prescriptive way in CAI, helping the student to focus on skills he needs to learn. Testing is not a rival explanation, but is another part of the variable being studied in "the CAI experience".

Instrumentation. Different instruments are used for the pretest O_1 and the posttest O_2. These tests are not equivalent tests that can be subtracted to determine the student's gain in reading skills. Instead, transformation methods will be used to standardize the pretest and posttest variables. Dometrius describes the process, "The standardizing process first identifies for a case the deviation above or below the mean of the variable ($Y - \bar{Y}$) and then transforms that distance to standard deviation units (divides by S_Y)." He explains, "As in other transformations discussed earlier this change does not affect what the variable measures; it just modifies the scale used for the measurement" (Dometrius, 1992, p. 423).

Interaction of selection and any other threat to internal validity. This threat considers that the group selected may react to one of the other threats to internal validity in a way different from other groups. This threat will be controlled as much as possible by using a large sample, randomly selected from the population. Also, the statistics will be replicated for developmental reading student populations over a period of four years.

The weakest part of this study is from the threats to internal validity. The limitation of the project is described by Fox, "Although we can unambiguously ascribe an observed difference to an experimental manipulation, we cannot unambiguously identify that manipulation with the independent variable that is the focus of our research" (Fox, 1997, p. 9). McMillan explains that internal validity "refers to the extent to which the independent variable, and not other extraneous or confounding variables, produced the observed effect (McMillan, 1996, p.194).

It is inappropriate to consider any one or two of these approaches as superior to the others. The effectiveness of a particular methodology depends in large part on the nature of the research question one wants to ask and the specific context within which the particular investigation is to take place. According to Dometrius, we need to gain insights into what goes on in education from as many perspectives as possible, and hence we need to construe research in broad rather than narrow terms (Dometrius, 1992, p. 11).

Generalizability:

Since the populations studied are approximately ninety-percent Hispanic, and are in an isolated area of Texas, results of this study can not be generalized to other colleges until the study is replicated on the other campuses. However, because the study has been replicated on this campus with eight developmental reading populations, some conclusions can be generalized to the next population of developmental students at UTB/TSC with some degree of success.

BEST COPY AVAILABLE

Only after the model is tested in many schools and with different populations can any generalizations be applied with other populations. Fraenkel describes how a research model becomes acceptable in other situations by repetition: "Such investigations, however, do not constitute science unless they are made public. This means that all aspects of the investigation are described in sufficient detail that the study can be repeated by any who question the results..." (Fraenkel, 1993, p. 6).

This paper will discuss the results from using multiple regression statistics to evaluate the success of CAI at UTB/TSC in Brownsville Texas. Suggestions will be made for replicating the model to evaluate CAI in other schools. Statistics and conclusions will be posted on my web-site at: http://unix.utb.edu/~lansford

References

Achen, Christopher H. (1982). Interpreting and Using Regression. Beverly Hills, CA: Sage Publications.

Boylan, Hunter R. (1998). An Evaluation of the Texas Academic Skills Program Austin, TX: National Center for Developmental Education, 1998, Under contract with the Texas Higher Education Coordinating Board (THECB), the NCDE conducted an extensive review of the Texas Academic Skills Program (TASP) from April through September of 1996, TASP Web site: www.tasp.nesinc.com.

Cohen, Elizabeth G. (1970). A New Approach to Applied Research: Race and Education. Columbus, OH: Charles E. Merrill Publishing Company.

Cohen, Jacob. & Cohen, Patricia. (1983). Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. Hillsdale: Lawrence Erlbaum Associates, Publishers.

Dometrius, Nelson C. (1992). Sociological Statistics Using SPSS. New York: Harper Collins Publishers Inc.

Fox, John. (1997). Applied Regression Analysis, Linear Models, and Related Methods. Thousand Oaks, CA: Sage Publications.

Fraenkel, Jack R. & Wallen, Norman E. (1993). How to Design and Evaluate Research in Education. New York: McGraw-Hill Inc..

Kidder, Louise H. (1981). Selltiz, Wrightsman and Cook's Research Methods in Social Relations. Dallas, TX: Holt, Rinehart and Winston.

Kleinbaum, David G. & Kupper, Lawrence L. (1978). Applied Regression Analysis and Other Multivariable Methods. Belmont, WA: Wadsworth Publishing Company, Inc.

McMillan, James H. (1996). Educational Research: Fundamentals for the Consumer. New York: Harper Collins College Publishers.

Pedhazur, Elazar J. (1982). Multiple Regression in Behavioral Research: Explanation and Prediction. New York: Holt, Rinehart and Winston.

Rudestam, Kjell Erik & Newton, Rae R. (1992). Surviving Your Dissertation: A Comprehensive Guide to Content and Process. Newbury Park, Sage Publications.

BEST COPY AVAILABLE